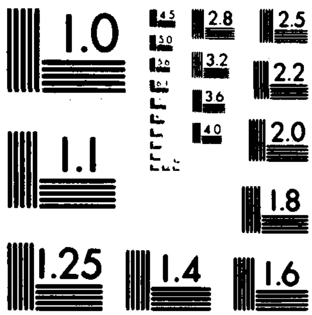
AD-A092 793 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
A SEQUENTIAL K-GROUP RANDOM ALLOCATION METHOD WITH APPLICATIONS--ETC(U)
NOV 80 A P SOMS DAAG29-80-C-0041
UNCLASSIFIED MRC-TR-624 NL

END
DATE
FILMED
1 81
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

LEVEL II (12)

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN

Madison, Wisconsin 53706

TECHNICAL REPORT NO. 624

November, 1980

A SEQUENTIAL k-GROUP RANDOM ALLOCATION METHOD WITH APPLICATIONS TO SIMULATION

by

Andrew P. Soms*

University of Wisconsin-Milwaukee

A Sequential k-Group Random Allocation Method with Applications to Simulation

Andrew P. Soms[*]

Abstract

A sequential method of random allocation is given and it is shown how it can be used to estimate the observed significance levels of k-sample nonparametric tests. The sequential technique is compared to the standard random allocation technique and shown to be more efficient. An application is made to the Dunn-Bonferroni method of multiple comparisons.

AMS(MOS) Subject Classification: 62L99, 65C05

Key Words: Dunn-Bonferroni method; Nonparametric tests; Observed significance levels; Simulation

Work Unit No. 4 - Statistics and Probability

---

[*]University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and the Office of Naval Research under Contract No. N00014-79-C-0321.

1. The Sequential Allocation Method

Bebbington (1975) showed that if there were N objects (such as file cards) from which it was desired to select (without replacement here and throughout) a random sample of size k without numbering the N objects, then one could proceed sequentially by selecting the first object with probability k/N and if at the $T^{\underline{th}}$ stage s have been selected, then the $T+1^{\underline{st}}$ object is selected with probability $(k-s)/(N-T)$, $T = 1,2,\ldots,N-1$.

We now state and prove the extension to an arbitrary number of groups. Suppose there are N objects and it is desired to sequentially divide them randomly into r groups of size $k_1,k_2,\ldots,k_r$, $\sum_{i=1}^{r} k_i = N$, i.e., each allocation has probability $1/\binom{N}{k_1,\ldots,k_r}$. Let $s_{1T},\ldots,s_{rT}$ be the number of objects selected for groups $1,2,\ldots,r$ at the $T^{\underline{th}}$ stage and let $P_{i,T+1}$ denote the selection probability for group i at the $T+1^{\underline{st}}$ stage. Then if

$$P_{i,T+1} = (k_i - s_{iT})/(N-T), \quad T = 0,1,\ldots,N-1 , \tag{1.1}$$

the selection is random. Note that $P_{i,1} = k_i/N$ and $\sum_{i=1}^{r} P_{i,T+1} = 1$. The randomness follows immediately by noting that the probability of a particular assignment is

$$\left(\prod_{i=1}^{r} k_i!\right)\Big/N! = 1\Big/\binom{N}{k_1,\ldots,k_r} .$$

Bebbington's (1975) result is a special case of the above when $r = 2$.

As an example, suppose $r = 3$, $k_1 = 2$, $k_2 = 2$, $k_3 = 3$ and $N = 7$. In order to make the sequential allocation given by (1.1) we take

7 independent random numbers U_i, $i = 1,2,\ldots,7$. Let

$$Q_{0,T} = 0 \text{ and } Q_{i,T} = \sum_{j=1}^{i} P_{j,T}, \quad i = 1,2,\ldots,r, \quad T = 1,2,\ldots,N.$$

Then the m^{th} object, $m = 1,2,\ldots,N$, is assigned to group n, where n is the unique integer such that

$$Q_{n-1,m} < U_m \leq Q_{n,m} .$$

Suppose the 7 random numbers are .79039, .01850, .99744, .81812, .93169, .22705, and .97709. The selection process is summarized in Table 1.

Insert Table 1 here.

Note that if all the k_i's are one, a random permutation is produced if we think of the group as denoting position.

2. Applications to Simulation

In k-sample nonparametric tests the observed significance level of the test is obtained by considering all possible partitions M of the (possibly tied) observed values or (possibly average) ranks into r groups, computing the value of the test statistic, and counting the number of times m it is equal to or greater than the observed value. The observed significance level $\hat{\alpha}$ is then m/M. When the number of partitions is large this is prohibitive and $\hat{\alpha}$ is estimated either by simulation (taking a large random sample of the allocations) or by asymptotics. The advantage of simulation is that one can control the accuracy of the estimate (by taking a large or small random sample) depending on the importance of the situation, unlike asymptotics which each time it is used forces one into the straight-jacket of committing a usually unknown error. Since it is (perhaps regretably) a well known fact that different actions will be taken for close values

of $\hat{\alpha}$, one above and the other below some fixed level (e.g., .01, .05, or .1) of the decision-maker, the use of simulation at least prevents approximating error in $\hat{\alpha}$ to be the determining factor.

If it is decided to use simulation, then a possible procedure is to make the random assignment as described in Section 1 many times by using a computer. The commonly used method is to produce a random permutation by ordering a random sample of uniform numbers and choosing the first k_1 indexes for group 1, the next k_2 for group 2, and so on. If all the k_i's are one, then this is more efficient than Section 1. However, as soon as the k_i's depart even moderately from 1, the method of Section 1 becomes much more efficient. As an example, if $k_1 = k_2 = k_3 = k_4 = 10$ and it is desired to make 2000 random assignments using a UNIVAC 1110 computer, a FORTRAN program using the methods of Section 1 uses 4.71 seconds of CPU time while a FORTRAN program using the random permutation method takes 9.17 seconds.

The Appendix contains a listing of the FORTRAN subroutine RANDM that uses the theory of Section 1 to make random assignments. This may be tied in with any specific simulation problem, e.g., the case treated in Section 3.

3. Applications to the Dunn-Bonferroni Method of Multiple Comparisons

The D-B (Dunn-Bonferroni) method is described in Dunn (1964). Briefly, let Y_{ij}, $i = 1,2,\ldots,r$, $j = 1,2,\ldots,n_i$, be continuous (this assumption is not important and is removed later) random variables with distribution function F_i, H_0: $F_1 = F_2 = \ldots = F_r$, H_a: for at least one pair (i,j), $F_i \neq F_j$ (in the sense of producing larger or smaller values), and the test must identify which, if any, pairs

are different. Denote by z_α the upper $\alpha^{\underline{th}}$ point of the standard normal. The D-B test declares all those pairs (i,j), $i < j$, different for which

$$Z_{ij} = |\bar{R}_i - \bar{R}_j| \Big/ \left[\frac{(N)(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)\right]^{1/2} \geq z_\alpha/(k(k-1)) \; , \quad (3.1)$$

where $\bar{R}_i$ denotes the average of the ranks of the $i^{\underline{th}}$ group in the joint ranking. The nominal significance level of this procedure is α. The actual significance level α_A is

$$\alpha_A = P_0\left(\max_{i<j} Z_{ij} \geq z_\alpha/(k(k-1))\right) \; , \quad (3.2)$$

and may be obtained by simulation based on Section 1. Table 2 gives some comparisons of nominal with actual, using Section 1 and 10,000 simulations.

Insert Table 2 here.

The Appendix contains a *listing* of the program used for Table 2. It thus appears that D-B is conservative and we can remove the conservatism by substituting for $z_\alpha/(k(k-1))$ $d_{(i)}$, where $d_{(i)}$, $i = 1,2,\ldots,r(r-1)/2$, is the $i^{\underline{th}}$ largest observed values of Z_{ij}, $i < j$, to obtain by simulation the $r(r-1)/2$ possible observed significance levels.

The K-S (Kruskal-Scheffé) method is also sometimes used in this situation (see, e.g., Miller, 1966, p. 166) and consists of replacing $z_\alpha/(r(r-1))$ in (3.1) with $h_\alpha^{1/2} = (\chi^2_{\alpha;r-1})^{1/2}$, where $\chi^2_{\alpha;r-1}$ is the upper $\alpha^{\underline{th}}$ point of χ^2 with r-1 degrees of freedom. The comparison of the critical constants in Table 3 shows that this is even more conservative than D-B.

Insert Table 3 here.

If the data is discrete, the D-B method can be modified as in Dunn (1964) and the random assignment done on average ranks. Thus ties present no problems in this approach.

The third method discussed in Miller (1964), the Steel many-one rank statistics, is too time-consuming for the simulation approach. For all practical purposes the exact D-B (use of the $d_{(i)}$ and simulation) seems the best method to use.

Appendix

```
C EXAMPLE OF MAIN PROGRAM FOR SEQUENTIAL RANDOM ALLOCATIONS
C NR IS THE NUMBER OF GROUPS, NSIM THE NUMBER OF SIMULATIONS, K'S THE GROUP
C SIZES, X'S THE NUMBERS TO BE ALLOCATED
      DIMENSION X(1000),Z(20,100),K(20)
  100 FORMAT ( )
   99 READ 100, NR,NSIM
      IF (NR.EQ. 0) GO TO 101
      READ 100,(K(J),J=1,NR)
      KK=0
      DO 1 J=1,NR
    1 KK=KK+K(J)
      DO 203 I=1,KK
  203 X(I)=I
      DO 2 II=1,NSIM
      CALL RANDOM(NR,K,KK,X,Z)
    2 CONTINUE
      GO TO 99
  101 STOP
      END
@FOR,IS ,SUB1
      SUBROUTINE RANDOM(NR,K,KK,X,Z)
C NR NUMBER OF GROUPS, K ARRAY OF GROUP SIZES, KK NUMBER OF ELEMENTS, X ARRAY
C OF KK ELEMENTS, Z RANDOM ALLOCATION OF X
      DIMENSION X(100),Z(20,100),NSC(20),Q(20),QC(20),U(1000),K(20)
      DO 3 I2=1,NR
    3 NSC(I2)=0
      DO 333 I2=1,KK
  333 U(I2)=RANDU(X)
      DO 4 I3=1,NR
      QC(I3)=0.
    4 Q(I3)=K(I3)
      MAXI=NR-1
      DO 5 II=1,MAXI
      DO 5 III=1,II
    5 QC(II)=QC(II)+Q(III)
      DO 6 I4=1,KK
      U(I4)=(KK-I4+1.)*U(I4)
      IF (U(I4).LE.Q(1)) GO TO 61
      DO 7 I5=2,MAXI
    7 IF (U(I4).LE.QC(I5)) GO TO 62
      IF (U(I4).GT.QC(NR-1)) INDEX=NR
      GO TO 64
   61 INDEX=1
      GO TO 64
   62 INDEX=I5
   64 NSC(INDEX)=NSC(INDEX)+1
      NN=NSC(INDEX)
      Z(INDEX,NN)=X(I4)
C UPDATE
      DO 8 I6=1,NR
      QC(I6)=0.
    8 Q(I6)= K(I6)-NSC(I6)
      DO 9 II=1,MAXI
      DO 9 III=1,II
    9 QC(II)=QC(II)+Q(III)
    6 CONTINUE
      RETURN
      END
```



```
C D-B BY SIMULATION
C NR IS THE NUMBER OF GROUPS, NSIM THE NUMBER OF SIMULATIONS, ZAL IS THE POINT
C FOR WHICH THE PROBABILITY OF THE MAXIMUM OF ALL THE ABSOLUTE VALUES OF
C STANDARDIZED RANK AVERAGE DIFFERENCES EQUALLING OR EXCEEDING IT IS TO BE
C CALCULATED, K'S ARE THE GROUP SIZES
      DIMENSION IFLAG(20,20),R(20)
      DIMENSION X(1000),Z(20,100),K(20)
  100 FORMAT ( )
   99 READ 100, NR,NSIM
      IF (NR.EQ. 0) GO TO 101
      MAXI=NR-1
      READ 100,ZAL
      READ 100,(K(J),J=1,NR)
      KK=0
      DO 1 J=1,NR
    1 KK=KK+K(J)
      DO 203 I=1,KK
  203 X(I)=I
      CON=KK*(KK+1)*ZAL**2
      COUNT=0.
      DO 2 II=1,NSIM
      CALL RANDOM(NR,K,KK,X,Z)
      DO 20 J1=1,NR
      R(J1)=0.
      NUPP=K(J1)
      DO 20 J2=1,NUPP
   20 R(J1)=R(J1)+Z(J1,J2)
      DO 21 J1=1,MAXI
      LLIM=J1+1
      DO 21 J2=LLIM,NR
      IFLAG(J1,J2)=0
   21 IF ((12*(K(J2)*R(J1)-K(J1)*R(J2))**2.GE.K(J1)*K(J2)*CON*(K(J1)+K(J2
     1)))) IFLAG(J1,J2)=1
      IPROD=0
      DO 22 J1=1,MAXI
      LLIM=J1+1
      DO 22 J2=LLIM,NR
   22 IPROD=IPROD+IFLAG(J1,J2)
      IF (IPROD.GE.1) COUNT=COUNT+1
    2 CONTINUE
      PR=COUNT/NSIM
      PRINT 100,(K(I),I=1,NR)
      PRINT 100,ZAL,NSIM,PR
      GO TO 99
  101 STOP
      END
```

References

Bebbington, A. C. (1975), "A Simple Method of Drawing a Sample Without Replacement", Applied Statistics, 24, 136.

Dunn, Olive J. (1964), "Multiple Comparisons Using Rank Sums", Technometrics, 6, 241-52.

Miller, Rupert J. (1966), Simultaneous Statistical Inference, New York: McGraw-Hill Book Company.

1. Selection Process

Stage	Random Digit	P_{1T}	P_{2T}	P_{3T}	Group Selected
1	.79039	2/7	2/7	3/7	3
2	.01850	2/6	2/6	2/6	1
3	.99744	1/5	2/5	2/5	3
4	.81812	1/4	2/4	1/4	3
5	.93169	1/3	2/3	0	2
6	.22705	1/2	1/2	0	1
7	.97709	0	1	0	2

2. Comparison of Actual to Nominal α

r	Common Group Size	Nominal α	Actual α
3	5	.05	.037
3	10	.05	.040
3	15	.05	.043
3	30	.05	.045
3	5	.01	.0030
3	10	.01	.0077
3	15	.01	.0077
3	30	.01	.010
5	5	.05	.026
5	10	.05	.036
5	5	.01	.0030
5	10	.01	.0067

3. Comparison of D-B and K-S Critical Constants

r	$z_{.05}/(r(r-1))$	$(\chi^2_{.05;r-1})^{1/2}$	$z_{.01}/(r(r-1))$	$(\chi^2_{.01;r-1})^{1/2}$
3	2.39	2.79	2.94	3.36
4	2.50	3.08	3.02	3.65
5	2.58	3.33	3.09	3.89
6	2.64	3.55	3.15	4.10
7	2.69	3.75	3.19	4.30
8	2.74	3.94	3.23	4.48
9	2.77	4.11	3.26	4.66
10	2.81	4.28	3.29	4.82

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR No. 624	2. GOVT ACCESSION NO.: AD-A092 793
3. RECIPIENT'S CATALOG NUMBER: 9 Technical rept.	
4. TITLE (and Subtitle): 6 A SEQUENTIAL k-GROUP RANDOM ALLOCATION METHOD WITH APPLICATIONS TO SIMULATION.	5. TYPE OF REPORT & PERIOD COVERED: Scientific - Interim
	6. PERFORMING ORG. REPORT NUMBER: 14 MRC-TR-624
7. AUTHOR(s): 10 Andrew P./Soms	8. CONTRACT OR GRANT NUMBER(s): 15 Army DAAG29-80-C-0041, [redacted] N00014-79-C-0321
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Dept. of Statistics & Math. Res. Ctr. University of Wisconsin Madison, WI 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research 800 N. Quincy Street Arlington, VA 22217	12. REPORT DATE: 11 Nov [redacted] 1980
	13. NUMBER OF PAGES: 12 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Dunn-Bonferroni method
Nonparametric tests
Observed significance levels
Simulation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A sequential method of random allocation is given and it is shown how it can be used to estimate the observed significance levels of k-sample nonparametric tests. The sequential technique is compared to the standard random allocation technique and shown to be more efficient. An application is made to the Dunn-Bonferroni method of multiple comparisons.

221 200